EXHIBIT V

Identifying Market Maker Trades as "Retail" from TAQ: No Shortage of False Negatives and False Positives

Robert Battalio
Professor of Finance
Mendoza College of Business
University of Notre Dame

Robert Jennings
Professor of Finance (Emeritus)
Kelley School of Business
Indiana University

Mehmet Saglam
Associate Professor of Finance
Carl Lindner College of Business
University of Cincinnati

Jun Wu
Research Support Director
Wharton Research Data Services

Abstract: Boehmer et al. (2021) propose a methodology to infer retail trades from publicly available NYSE Trade and Quote (TAQ) data. Their methodology relies on assumptions about what types of orders do and do not trade on non-quote-midpoint sub-penny increments via the Trade Reporting Facility (TRF). We obtain proprietary data from one or more wholesalers known to receive marketable orders from retail brokers. We use these data to demonstrate that the Boehmer et al. (2021) methodology identifies less than one-third of trades generally assumed to be from retail investors and analyze cross-sectional determinants of the technique's identification rate. In addition, we obtain proprietary data on institutional trades from multiple sources and demonstrate that a large number of such trades print on the TRF at non-quote-midpoint sub-penny prices in violation of the assumption that institutional orders trade only on penny or half-penny increments. Thus, there are both Type I and Type II errors that affect the ability to identify retail and only retail trades from TAQ using the Boehmer et al. (2021) methodology. Finally, we demonstrate that these errors can produce different inferences regarding the association between lagged retail order imbalance measures and stock returns.

Current draft, September 21, 2023.

Acknowledgments: The authors received no outside funding for this study and have no conflicts of interests related to this study. The authors thank John Shim and participants in the University of Maryland finance department's Workshop and The Microstructure Exchange for helpful comments on previous versions of the paper. Finally, we thank one or more wholesalers for providing some of the data used in this paper.

A considerable body of academic research investigates the trading behavior of retail investors. Key to all such pursuits is identifying the trades of retail traders. One line of research utilizes proprietary data to study a subset of retail investors. The seminal paper by Odean (1998) generated a plethora of papers that utilize data from a discount stock broker to examine retail trading behavior. Kaniel et al. (2008) utilize a proprietary dataset provided by the New York Stock Exchange (NYSE) to study individual investor trading practices. Since the passage of Regulation NMS, which greatly reduced the NYSE's market share, it is difficult to characterize retail investor behavior using data from a single stock exchange. Kelley and Tetlock (2013) address this difficulty by obtaining a proprietary dataset that includes retail orders executed by an order flow wholesaler[1] between 2003 and 2007 to examine retail trading behavior. Unfortunately, retail brokers or the execution venues to which they send their orders rarely provide proprietary transaction level data.

A second approach uses algorithms intended to identify retail trades in the NYSE's Trade and Quote (TAQ) database, which has been publicly available since 1993. One of the first algorithms used to identify retail trades from TAQ data was trade size. Reilly (1979) asserts that trades of 1,000 or fewer shares are made primarily by individuals and uses this cutoff to identify institutional trades when analyzing transaction records obtained from Francis Emory Fitch, Inc. Lee (1992) uses a dollar-based threshold of $10,000 to distinguish between retail and institutional trades. Once identified as a retail trade, the marketable side of the transaction is typically inferred using some version of the Lee and Ready (1991) algorithm. Lee and Radhakrishna (2000) use the NYSE TORQ database to demonstrate that trade size could be effective at distinguishing between

[1] A wholesaler is an entity that obtains order flow from multiple retail brokers (frequently with the brokers charging the wholesaler payment for order flow) and executes those trades either by taking the other side from their inventory (internalization) or through connections with other execution venues such as exchanges and Alternative Trading Systems (e.g., dark pools). The wholesaler typically provides price improvement relative to National Best Bid or Offer quoted prices and fills orders with sizes larger than quoted depth at quoted prices or better. The wholesaler relieves the broker of having to build and maintain the competitive connectivity to multiple execution venues necessary to satisfy Regulation NMS.

retail and institutional traders. Cready, et al. (2014), however, note that a significant concern in studies using trade size cutoffs to identify retail investors "is spurious effects attributable to misclassification of transactions, particularly those originating from large investors." Hvidkjaer (2008) posits that the likelihood that a small trade is a piece of a large institutional order became much higher after the shift to decimal pricing in 2001 and Reg NMS in 2005. As a result, most researchers currently avoid using trade size cutoffs to identify retail and institutional trades in U.S. equity markets.

Boehmer, Jones, Zhang, and Zhang (2021), hereafter BJZZ, propose and develop a methodology to infer retail purchases and sales from publicly available data by identifying trades with sub-penny prices that are reported to FINRA's Trade Reporting Facility (TRF) rather than through an exchange. BJZZ assert that "in the United States, most marketable equity orders initiated by retail investors do not take place on one of the dozen or so registered exchanges." Off-exchange trading volume represents a significant portion of total consolidated volume. In May 2022, Rosenblatt finds that off-exchange trading was 40% of total U.S. equity trading volume. Further, Rosenblatt estimates that wholesalers account for 42.5% of off-exchange trading, dark pools about 23%, and capital commitment and manual crossing of institution trading interests another 22%.[2] To separate retail and institutional trades reported to the TRF, BJZZ's methodology assumes that only retail trades print on sub-penny prices at any increment other than a price equal to the midpoint of the quoted spread. As noted in BJZZ, most retail order flow is routed to so-called order flow wholesalers (often, but not always with the retail brokers charging payment for order flow) and provided price improvement as the wholesalers compete to provide best execution for their clients' orders (see, e.g., Battalio and Jennings (2022)). Conversely, BJZZ state that

[2] See Trading Talk – Market Structure Analysis: An Update on Retail Market Share in US Equities, June 24, 2022.

institutional trades generally cannot receive non-midpoint prices but that institutional trades "often occur at the midpoint of the prevailing bid and ask prices." BJZZ assert that their "approach is therefore likely to pick up a majority of the overall retail trading activity."

Using proprietary retail and institutional order/trade data provided by one or more wholesalers[3], institutional orders executed by a major investment bank, and order data from a public pension fund, we examine the accuracy of the BJZZ algorithm in identifying retail trades and excluding institutional trades in the NYSE's TAQ database. We find that their procedure identifies less than one-third of trades known to be retail and frequently could include known institutional trades as retail. Furthermore, we demonstrate that these identification errors result in differences in conclusions regarding the association between order imbalance metrics and security returns.

The methodology in the BJZZ paper is widely employed to infer retail trading and to draw conclusions about retail trading.[4] Within several of these papers, however, are caveats regarding the ability of the BJZZ algorithm to correctly identify retail trades in the publicly available TAQ database. For example, Blankespoor et al. (2018) examines the market's response when the Associated Press began using algorithms to write articles about firms' earnings announcements. Using the BJZZ algorithm to identify retail trades in TAQ, the authors present evidence that retail trading increases around the release of these articles. The authors suggest in a footnote that this conclusion relies, in part, on the assumption that by excluding trades that execute at the round penny or around the half-penny they are eliminating institutional trades. Using the BJZZ algorithm to identify retail trades in TAQ, Barber et al. (2021) present evidence that Robinhood investors

[3] For literary convenience and brevity, we refer to "one or more wholesale(s)" with simply wholesaler(s) in the remainder of the paper when referring the data providing wholesaler(s).
[4] See, for example, Bonsall, Green, and Muller (2020), Bushee, Cedergren, and Michaels (2020), Farrell et al (2022), Guest (2021), and Israeli, Kasnik, and Sridharan (2021). As of the date of this draft, Google Scholar lists 298 citations.

engage in more attention induced trading than other retail investors. In a footnote, however, they state that "this conclusion assumes that there is no bias in the Boehmer et al. (2021) methodology that would affect concentration measures." Bradley et al. (2022) write that the BJZZ methodology "is conservative in the sense that it has a low type I error (i.e., trades classified as retail are very likely to be retail)" but "does omit retail trades that occur on exchanges."

In this draft of the paper, we analyze the ability of the BJZZ methodology to identify known retail trades executed by the cooperating wholesaler(s) during the first ten trading days of December 2020 and find that it identifies fewer than one in three of these retail trades. We will expand the sample period of wholesaler(s) retail trading data in future drafts of the paper. In addition, we look at three samples of non-retail trade executions that print in non-midpoint sub-penny increments. Our data sources span electronic liquidity providers' single-dealer platforms, a major investment bank's dark pool, and a buy-side firm's usage of several dark pools and single-dealer platforms. We first provide summary statistics on a sample of over two million such institutional trades from our retail-order data provider(s) during December 2021. Secondly, we analyze a sample of large institutional orders with observed individual child trades executed by a major investment between January 2011 and March 2012.[5] Finally, we examine a sample of trades with multiple dark pools and electronic liquidity providers from the Colorado Public Employees Retirement Association for 2016 and 2017 as a way to demonstrate that our institutional trade results are not specific to particular execution platforms.[6] Overall, we document the potential for many Type I errors.

It is well-documented in the literature that institutional order flow can be informative (e.g., Hendershott, et al. (2015)). One of the assumptions used by the BJZZ algorithm is that institutional

[5] These are the same data used by Battalio, Hatch, and Saglam (2022).
[6] Data obtained via a Freedom of Information Act filing associated with a different research project.

order flow executed in Alternative Trading Systems (e.g., dark pools) and non-ATS execution venues (e.g. single-dealer platforms) and reported through the TRF does not receive non-midpoint sub-penny price improvement. Evidence to the contrary would make it difficult to interpret the results of studies that examine whether “retail” trades identified using the BJZZ methodology are informed. Our data-providing wholesaler(s) identify 19,802,471 institutional trades in December 2021 and, after eliminating 2,741,318 trades that receive midpoint pricing, furnishes us a sample of over two million (almost 11% of total trades) non-retail trades filled at non-midpoint sub-penny prices on venues that BJZZ argue do not receive such prints. We also use an alternative set of institutional trades from a major investment bank to examine whether this assertion by BJZZ is correct. Of the 166,266 sample institutional trades that obtain liquidity from electronic liquidity providers like Citadel Securities, Getco, and Knight between January 2011 and March 2012, we find that over 78% would be classified as retail trades by BJZZ. Looking separately at the 136,832 institutional orders executed in the broker’s ATS, one-third of these trades are classified as retail by the BJZZ algorithm. Thus, a substantial portion of institutional trades with electronic liquidity providers or in an investment banking firm’s ATS could be misidentified as retail trades. Finally, we examine the trades of a pension fund across multiple execution venues. Of the 363,459 (6,203) pension fund’s trades in dark pools (with electronic liquidity providers), we find that about 6.1% (26.4%) would be classified as retail using BJZZ. Whether or not these Type I errors are severe enough to alter inferences in all studies of retail investor trading behavior is a question that we cannot answer here. However, we do provide evidence consistent with the claim that this type of error can affect inferences about the association between lagged measures of retail order imbalance and stock returns.

We also obtain all marketable retail orders routed to the cooperating wholesaler(s) during our sample period. Interestingly, and in stark contrast to the assertion made in BJZZ that retail orders are seldom routed to exchanges, 19.6% of the trades generated by the retail orders that are routed to the wholesaler(s) execute using liquidity sourced outside the wholesaler(s) (e.g., exchanges or other sources of liquidity) and thus interact with other (i.e. non-data-provider(s)) order flow. We match the proprietary trades that are known to be retail to trades reported to the TRF obtained from the NYSE's TAQ database. We use this sample of TAQ trades to evaluate the likelihood of Type II errors (i.e., that the BJZZ methodology fails to identify them as retail trades). Nearly 40% of the matched trades have execution prices that have no sub-penny prices and roughly 30% of the matched trades have trades that have a sub-penny price in the interval [0.4, 0.6] and, thus, are not identified as retail trades by BJZZ. As a result, only about 30% of the sample of retail trades that are matched to TRF trades are classified as retail by the BJZZ methodology. Just over 94% of these trades have the correct inferred order side using the BJZZ methodology.

Do Type I errors (i.e., identifying non-retail trades as retail trades) and Type II errors (i.e., the failure to correctly identify retail trades) result in any substantive differences in the inferences researchers using the BJZZ methodology have made? We investigate one such inference. BJZZ examine the association between order imbalances constructed using inferred retail trades and future stock price movements to determine whether retail order flow appears to be informed. Our proprietary data provider(s) also produces the four BJZZ-defined order imbalance measures on a stock-day basis using all of their retail trades for the period of time from August 3, 2020 through July 26, 2022. We construct analogous order imbalance measures for BJZZ-identified trades for these stocks from TAQ during the same period and compare them to the order imbalance measures from the data provider(s). We find that the correlations between the actual and the BJZZ-inferred

order imbalance measures are less than one-half as high as BJZZ found when comparing their order imbalance measures to order imbalance (OIB) measures created using a sample of proprietary retail trades provided by Nasdaq. Furthermore, we find systematic differences between the wholesaler(s) order imbalance measures and the BJZZ-inferred metrics. Specifically, the BJZZ-inferred OIB measures are less sell/more buy oriented than the "true" OIB measures using the data-provider(s) retail orders.

Using the two sets of order imbalance measures to investigate their association with future stock returns, we find that the wholesaler(s) measures and the BJZZ measures produce similar results when examining the subset of securities that were the focus of the BJZZ analysis. The wholesaler(s) order imbalance measures actually are somewhat more strongly correlated with future returns than are the BJZZ measures. In an attempt to distinguish between the effects of Type I and Type II errors, we divide the sample securities into quintiles based on the fraction of share volume that the six major wholesalers represent of the total number of shares traded on the consolidated tape (i.e., TAQ). Low retail intensity quintiles presumably suffer from proportionally more Type I error and high retail intensity quintiles arguably suffer relatively less Type I error. Results from the low retail intensity quartiles indicate that the wholesaler all-trade and all-volume order imbalances (unaffected by either Type I or Type II errors) are significantly positively correlated with future stock returns while BJZZ-inferred order imbalance measures are not strongly associated. This suggests that inference errors affect the ability of the BJZZ methodology to detect at least some statistical associations. Using only high retail intensity securities, we find that both the BJZZ order imbalance measures and, to some extent, the wholesaler(s) order imbalance measures are statistically significantly correlated with future returns for the all-trade and the odd-lot OIBs. When focusing on just the common stock sample from BJZZ, there is less

disagreement between the wholesaler(s) OIB regressions and the BJZZ-inferred OIB regressions; for low retail intensity stocks neither set of OIB metrics is correlated with future returns but with high retail intensity stocks both are positively correlated with future returns. These findings suggest that both Type I errors and the sample of securities examined can influence the inferences made when using BJZZ-identified retail trades to predict future returns.

In the next section, we discuss the differences between our paper and two closely-related papers. In section III, we introduce our three samples of proprietary institutional trade data and demonstrate that these trades frequently do print on non-midpoint sub-penny increments. We then turn to evaluating our proprietary retail order data set for Type II errors in Section IV and find that only about 30% of the known retail trades are properly identified as such by the BJZZ methodology. In Section V, we replicate the methodology of BJZZ in identifying a statistical association between lagged order imbalances and returns and demonstrate that we obtain differing results between our proprietary retail trades (virtually free from either Type I or Type II errors) and BJZZ-inferred portfolios of securities with varying proportions of the two types of inference errors. Lastly, we conclude.

II. Literature Review of Closely-Related Research.

Barber et al. (2022) conduct an experiment to evaluate the effectiveness of the BJZZ methodology at identifying retail trades by placing over 85,000 orders in 128 stocks between December 21, 2021 and June 9, 2022 through six retail brokers. The 128 sample stocks are the result of a randomized sampling process after stratifying all stocks with a CRSP security code of 10 or 11 priced greater than $1.00 on market capitalization, liquidity (turnover), volatility, and stock price. In their main analysis, they submit orders for a $100 notional amount (requiring an integer number of shares) or, if the share price exceeds $100 one share between 9:40am and

3:50pm.[7] To mimic day trading activity, they initiate positions by buying and then selling 30 minutes later and carry no inventory overnight. They find that the BJZZ methodology identifies about 35% of their actual retail trades as retail and correctly signs (as buys or sells) about 72% of those. Furthermore, they conclude that, on a stock basis, 30% of BJZZ-constructed order imbalance measures are uninformative because the accuracy rate for signing trades does not differ from 50%.

We can distinguish our paper from Barber et al. (2022) in several dimensions. We are the first, to our knowledge, to demonstrate that institutional trades reported on the TRF can print in sub-penny intervals other than the half-penny. Thus, the notion that BJZZ identify as retail trades only trades that are from retail investors is incorrect.

Second, our sample of known retail trades is much different from Barber et al. (2022). Our sample of retail orders reflects overall retail trading interests – it contains orders from almost 9,600 trading symbols, many not CRSP security codes 10 or 11. Retail investors trade a number of securities (e.g., Exchange Traded Funds) that are not simple common stocks.[8] It is important to preview our sample of BJZZ-matched actual retail orders with their sample. On a trade-weighted basis, our data's mean trade price is $123.94 and the mean trade size is 231 shares. This implies an average notional trade size of $28,630, much larger than the $118 average trade size in their sample. We demonstrate that the success of the BJZZ methodology is sensitive to order and trade size as well as stock price. Our order-weighted mean National Best Bid Offer spread is $0.09 versus their reported mean spread of $0.17. Less than 20% of their trades occur when the NBBO

[7] In a robustness check Barber et al. (2022) submit 2,292 orders for a notional amount of $1,000.
[8] In Section V of this paper, we rank each security based on its retail intensity defined as the marketable order executed volume of the six major wholesalers in that symbol divided by the total trading volume in that symbol from TAQ. If we impose the BJZZ restrictions that the symbol represents a common stock and be listed on the NYSE, the NYSE MKT, or Nasdaq, we eliminate over 86% of the top two retail-intensity quintile symbols.

quoted spread is $0.01 but about 44% actual retail trades identified by the BJZZ methodology in our sample occur when the spread is at minimum. This difference might be at least partially explained by the fact that liquidity tends to concentrate at the open and the close – almost 10% of our sample trades occur before 9:40am or after 3:50pm.

Barardehi, Bernhardt, Da, and Warachka (2022), hereafter BBDW, argue that wholesalers internalize retail order flow to manage inventory imbalances created when supplying institutional investors liquidity via the wholesalers' market making on exchanges, participation in Alternative Trading Systems, and/or trade in the wholesalers' single-dealer platforms. When institutions demand one-sided liquidity, wholesalers balance that by internalizing opposite-sided orders from retail order flow, which are reported on the TRF frequently as trades executing at non-midpoint sub-penny prices. Thus, BBDW argue that the BJZZ algorithm is inherently designed to identify the portion of wholesaler retail trading most closely associated with *institutional* liquidity demand. As a result, retail order imbalances are "…informative about the consumption of liquidity by institutional investors precisely because the BJZZ algorithm identifies a key subset of retail orders." BBDW demonstrate that *intraday* returns are negatively correlated with BJZZ retail order imbalance measures but positively correlated with institutional order imbalances. The subsequent unwinding of the institutional price pressure over longer periods of time is the driving force that underlies the positive return predictability documented by BJZZ, not the information content of retail trading.

One conclusion of BBDW is that the retail order imbalance association with returns relies critically on BJZZ identifying only the internalized order flow that is designed to offset institution order imbalances. Using the BJZZ order imbalance variable name Mroib, they state that

> "As our model illustrates, if one were to include in Mroib the retail orders that a wholesaler executed on a riskless principal basis (differentially on the same side as

institutional demand), it would reduce the information content regarding institutional liquidity demand."

We, however, find that order imbalance measures including all of our data-providing wholesaler(s)' retail order flow (both internalized and externalized trade) produces highly significant statistical associations with weekly returns as strong or stronger than BJZZ imputed order imbalance measures. In addition, in a companion paper Battalio and Jennings (2022), we demonstrate that even when the wholesaler(s) externalize orders they frequently do not simply accept the price received from the external execution venue (i.e., conduct riskless principal trades) but provide price improvement to the customer at the wholesaler's expense. This price improvement to externalized trades counteracts the cost effects of internalized trades in the BBDW model.

III. Type I Errors.

To better understand whether trades generated by institutional orders executed away from exchanges are as BJZZ state on their page 2,255 "usually in round pennies," we obtain three samples of non-retail orders. As discussed previously, the institutional investor order/trade data used to examine Type I errors (institutional trades falsely identified as retail trades) come from cooperating wholesaler(s), a major investment bank, and a public pension fund.

Firstly, we obtain a sample of 2,100,769 institutional trades that occur at sub-penny but not half-penny prices during December 2021 from our data provider(s). These trades represent the wholesaler(s) providing liquidity to institutional orders at sub-penny prices other than a half cent and reported to the TRF. Of the 19,802,471 institutional trades examined, 4,842,087 (24.45%) occur at sub-penny prices. Of those, 43.4% occur at non-half-penny sub-penny intervals (i.e., 2,741,318 occurred at the half-penny). Of the 2,100,769 trades occurring at non-half-penny sub-penny increments, 2,100,162 occur during regular trading hours and comprise the sample we

analyze. The majority of these trades (52.38%) are sell orders. Almost 92% of these trades execute on the non-ATS single-dealer platform(s) operated by the wholesaler(s) (in a single-dealer platform, the operator is the sole counterparty) and the remainder are roughly evenly split between Alternative Trading Systems (dark pools and Electronic Communication Networks) and Exchange Retail Liquidity Programs. (The latter, although representing a retail trade, is a retail trade occurring on and reported through an exchange.) We provide descriptive statistics on these trades in Table 1.

[Insert Table 1 about here.]

Although these trades do not result from retail orders, it would be difficult to distinguish them from retail trades based on order size, trade price, or trade time statistics detailed in Battalio and Jennings (2022) and in the next section of this paper (see Table 6). The mean trade size is 302 shares and over 86% of the institutional trades are for fewer than 500 shares.

[Insert Table 2 about here.]

In Table 2, we summarize the frequency of various sub-penny pricing intervals for this set of institutional trades. Of the 2.1 million institutional, non-half-penny, sub-penny trades, only 17% would be eliminated using BJZZ's "near one-half penny" exclusion rule (recall the data provider(s) removed 2,741,318 half-penny sub-penny trades). Thus, nearly 1.75 million institutional trades (about 9% of the total) from our data provider(s) would be included as retail using BJZZ's algorithm.

We also obtain a sample of institutional parent orders and the corresponding child order executions processed by a large investment bank's (IB's) volume weighted average price algorithm.[9] This data set was also used by Battalio et al. (2022) and cover large institutional orders

[9] Volume weighted average price algorithms are one of the most commonly used trading algorithms that seek to match the volume-weighted average price realized during the trading horizon.

in S&P 500 stocks between January 2011 and March 2012. Here, we consider the entire set of 166,266 child order executions in the IB's dark pool and 136,833 child order executions that source liquidity from electronic liquidity providers like Getco, Citadel Securities, and Knight Securities. Table 3 contains the sub-penny pricing distribution for each collection of trades.

[Insert Table 3 about here.]

Focusing first on the second column of Table 3, we see that 38.4% of the trades filled in the IB's dark pool executed in round pennies and 28.4% of the trades executed with sub-penny increments in the range [0.4, 0.6]. This implies that 33.2% of the institutional trades executed in the IB's dark pool have sub-penny pricing increments, which are classified as retail by the BJZZ algorithm. Moving to the third column of Table 3, we see that 78.3% of the trades executed by ELPs away from exchanges have sub-penny pricing increments that make the trades eligible to be classified as retail by the BJZZ algorithm.

[Insert Table 4 about here.]

Thirdly, we employ the dark pool and electronic liquidity provider trades of the Public Employee's Pension Association of Colorado during 2016-2017. Table 4 provides an analysis of the sub-penny trade prices of COPERA. Panels A and B contain dark pools. Panel A reports the results of trades in the dark pools operated by the broker COPERA used for the parent orders and Panel B dark pool trades not operated by the broker handling COPERA's order. Panel C reports the results of trades with electronic liquidity providers (again, such as wholesalers). In the broker's own dark pool trades about 15.6% of the trades occur at sub-penny prices BJZZ would classify as retail trades. In dark pools not operated by the broker handling the order, the percentage of BJZZ-inferred retail trades is much smaller, approximately 3.8%. Using ELPs, 26.4% of the pension fund's trades potentially are classified as retail by BJZZ.

The results from all three sets of institutional orders contradict the assertion that institutional trades executed in the dark "are usually in round pennies" and suggests that Type I errors may plague studies that use BJZZ-identified trades to examine ***retail*** trading behavior. There is the potential for these magnitudes of mis-identification to produce many incorrect retail attributions. As noted above, Rosenblatt estimates that wholesaler (ELPs) account for about 17% of consolidated tape trades and dark pools approximately another 9% during May 2022. In that month, TAQ report 1,919,062,354 trades in the consolidated tape. Across our three samples of ELP trades the estimates of non-half-penny sub-penny prints range from 11% to 78% and across our two samples of dark pool trades the analogous range is 6% to 33%. At the low (high) end of the range, this implies 46.3 (311.5) million potentially mis-identified trades per month. The fact that Reg NMS prohibits "orders from having sub-penny limit prices" does not appear to restrict institutional trades from being executed with non-midpoint sub-penny price increments and, therefore, be identified as a retail trade.[10]

IV. Type II Errors.

A. Data.

To begin our analysis of Type II errors in the BJZZ methodology, we obtain proprietary marketable order and trade data from the cooperating wholesaler(s) for the month of December 2020. We receive all of the marketable retail orders handled by the wholesaler(s) during this time period. The order data include: date and time of order entry, a unique order identification number, stock trading symbol, the type of order (market or marketable limit), the limit price if applicable, the order's side (buy or sell, with an indicator variable for short sell), and the order quantity. The trade data include: date and report time of trade, the unique order identification number mapping

[10] BJZZ write that "in the early part of our sample, a small number of dark pools allowed some sub-penny orders and provided non-midpoint sub-penny execution prices, but our results hold when we exclude this subperiod."

back to the order data, a unique trade reference number, the stock trading symbol, the number of shares filled by this execution, and the execution price. Using the order identification number, we can merge order and trade data.

We restrict our analysis of Type II errors to trades reported via FINRA's Trade Reporting Facility (TRF) and not to an exchange. In order to facilitate potential analyses by stock-day, we restrict our sample to stock symbols averaging at least 100 round-lot trades per day (2,200 trades for the sample month) without any days reporting zero TRF trades. Furthermore, we require that the end-of-month stock price be greater than one dollar to mitigate sub-penny limit order pricing. That provides us with a sample of 2,741 stock symbols (slightly less than 29% of the symbols in the original database). The sample stocks produce 85% of the trades contained in the proprietary data. We provide some descriptive statistics of our sample in Table 1.

[Insert Table 5 about here.]

End-of-month share price ranges from the designed minimum of $1.01 to over $3,000 and averages $64.57. Round-lot trades for the month range from the designed minimum of 2,201 to nearly two million and averages 23,328. Overall, there is substantial variation in the two variables used to restrict the sample stocks.

The BJZZ methodology uses publicly available TAQ data to infer retail trades and order sides. To assess the success of this methodology for identifying a sample of known retail trades, we match our set of proprietary retail trades to the corresponding TAQ trades (for which we will use the BJZZ methodology to infer its retail status). To begin this process, we gather all trades reported to the TRF (exchange code 'D') in the restricted-sample stocks for December 2020 from Daily TAQ. During the sample month, there were 327,542,261 trades for the sample stocks

reported via the TRF. The mean trade price was $109.84 with a range of $0.30 to $3248.99.[11] There are a disproportionate number of trades in higher priced stocks as the trade-weighted trade price exceeds the equally-weighted share price. Mean trade size was 240 shares with a range of one share to 6,199,125 shares. Following BJZZ, we exclude trades with non-normal condition codes and trades executed at prices less than $1 in the analysis below.

In this version of the paper, we restrict our analysis to the first ten trading days of December 2020. Future drafts will expand the sample period. We match the proprietary data trades to TAQ TRF trades based on symbol, time, price and quantity for the 2,741 activity-restricted-sample securities. Symbol, price and quantity are unambiguous matching criteria. We have two times; one from the data provider(s) and another from TAQ. We choose TAQ's Participant Timestamp as the benchmark (the time the participant reported the trade to the SIP) because it is the earliest TAQ timestamp and allow a maximum of ten milliseconds difference between it and the proprietary data timestamp.[12] We match 78.16% of the proprietary trades with 81.25% of the volume to TAQ trades within the ten-millisecond window. Examining the trade matching success on a stock-level basis, the mean (median) matching success is 75.05% (75.92%) with a range from 22.21% to 95.88%. Given that nearly 20% of the wholesaler(s)' trades execute using liquidity sources external to the wholesaler(s) (e.g., exchanges and ATSs), we did not expect to match all of the data-provider(s) trades as these externalized trades more frequently involve complex order routing strategies that utilize exchanges and longer elapsed times between initial order arrival time and eventual trade

[11] Clearly, requiring the stock to have a trade price greater than $1 at month's end was insufficient to eliminate all trade prices less than $1. The BJZZ methodology we replicate later in the paper eliminates all trades priced less than $1.

[12] Participant timestamp in TAQ is measured in microseconds for exchange traded stocks instead of nanoseconds for the data provider(s).

execution. Table 6 provides some descriptive statistics regarding the trades we were and were not able to match.

[Insert Table 6.]

Overall, the retail trades that we match to TAQ appear to be a representative sample of the original retail order data set. The matched trade sample is somewhat more likely to be a buy order and be in a stock with a higher transaction price. The mean order size for the matched orders is the only order/trade characteristic for which the matched sample differs remarkably from the original proprietary data set. The mean order size is substantially smaller for the TAQ-matched trades than the overall retail database. Not surprisingly, the larger orders are less frequently completely filled in a single trade with wholesaler(s) internalization so it is likely that proportionally fewer of these trades make their way to the TRF (as noted previously, 19.6% of the trades are consummated by the data provider(s) accessing external liquidity sources from venues including exchanges). Finally, the mean time difference between the proprietary data timestamp and the TAQ Participant Time is just over one millisecond (result not tabulated in Table 6).

B. Failing to identify retail trades as retail trades.

We use the BJZZ methodology to infer which of the TAQ-matched trades came from retail investors (all of which are presumed to be retail trades). To replicate BJZZ, we compute their classification variable $Z_{i,t}$ = 100 * mod($Price_{i,t}$ .01), where the subscript i,t denotes stock i at time t. Table 3 contains the sub-penny pricing distribution for matched retail trades by trade side conditional on the order-receipt time NBBO width.

[Insert Table 7 about here.]

We find that 39.66% of the matched retail trades have no sub-penny pricing and 29.90% have sub-penning pricing that BJZZ consider as midpoint pricing (and therefore not classified as retail). Of trades with sub-penny increments in the range [0.4, 0.6], 91.74% of those trades are

exactly at a sub-penny pricing increment of 0.5 (not tabulated on Table 7). Thus, only 30.44% of our sample of matched retail trades has $Z_{i,t}$ in the (0, .4) or (.6, 1) intervals and is classified as retail by the BJZZ methodology. This is slightly lower than the Barber et al. (2022) identification rate of 35%. When the quoted spread is $0.01, the BJZZ algorithm does slightly better identifying just over 34% of the known retail trades as retail.

[Insert Table 8 about here.]

In Table 8, we provide some descriptive statistics regarding our experience using the BJZZ methodology to classify our activity-restricted-sample of known retail trades as retail. From Panel A, we determine that the BJZZ methodology's inferred retail trade sample differs from the known retail trade universe in that the average trade (order) size is somewhat larger (smaller), the average trade-weighted trade price is slightly lower, and the average order receipt time is slightly later in the trading day. Panels B and C demonstrate that the methodology is least effective at identifying known retail orders throughout the opening half hour of trading. Interestingly, Panels D and E suggest that BJZZ's algorithm is less effective at identifying odd lot orders and trades than round or partial-round orders and trades.[13] Panel F finds that the BJZZ approach is more effective at identifying retail trades for stocks with prices less than $100 per share, likely because these stocks are more frequently quoted at the minimum tick size of $0.01.[14]

We now turn to examining the success of BJZZ's methodology in correctly inferring trade side. For this analysis, we focus on the 7,349,520 proprietary data trades the BJZZ methodology

[13] This is consistent with Barber et al (2022), which finds higher identification rates for 100 share orders than odd lots. Bartlett (2022) suggests that is due to SEC Rule 605, which requires venues to post execution quality statistics for round lots but not for odd lots, however, evidence in Battalio and Jennings (2022) finds extensive price improvement for odd lots.

[14] Finally, in results that are not tabulated in this version of the paper, we examine orders filled in a single execution versus orders filled in multiple executions. We find that about 93.6% of the retail orders are filled in a single trade. 99% of orders have four or fewer trades but the maximum number of trades in an order can be quite large (maximum of 753). For multiple-fill retail orders, BJZZ identify all trades associated with the order as retail for only about one of five orders.

identifies as retail trades. Overall, we find that just over 94% (compared to their check with limited Nasdaq DarkRetail data of 98.2% for stocks with prices less than $100) of BJZZ-identified retail trades have the correct inferred side.[15] This suggests that about 6% of the BJZZ-identified retail trades are getting substantial (i.e., better than midpoint) price improvement, which results in misclassification of order side by BJZZ.[16] In Table 9, we provide some detail on the trade side inference of BJZZ's approach.

[Insert Table 9 about here.]

From Panel A we find that it is about 1.5 times more likely for BJZZ to misclassify a buy order as a sell than it is to misclassify a sell order as a buy (3.59% versus 2.29%).[17] In Panel B, we examine BJZZ's side inference success by time of day. The BJZZ algorithm is slightly less successful in the first half hour of trading but there is not substantive variation in success rate across the day. In Panel C, we examine BJZZ's order side inference success rate by order size and, more practically, trade size. Generally speaking, the methodology is slightly less accurate for smaller orders and trades than for large orders/trades suggesting that small orders/trades are more likely to receive greater price improvement from the wholesaler(s) than larger orders/trades. In Panel D, we examine the algorithm's success in identifying whether a retail trade is buyer- or seller- initiated conditional on whether the retail order generating the trade was filled with a single execution. For the 5,971,948 orders filled with a single execution that are identified as retail by the algorithm, the inferred order side is the actual order side 94.85% of the time. For retail orders filled with multiple trades, the overall success rate for inferring order side is less than 63%. The algorithm correctly infers the order side for all of the trades generated by an order requiring

[15] This differs markedly from Barber et al (2022) likely due to a difference in sample stock selection.
[16] Note that executions on a full penny price might represent price improvement when quoted spreads exceed $0.01.
[17] The magnitude and bias in misidentification of order side with our data are greater than and opposite that documented in BJZZ (page 2261).

multiple trades only about 36% of the time (258,546 of 711,641 BJZZ-identified retail orders with multiple fills). For a relatively small number of orders with multiple trades (65,905 of BJZZ-identified retail orders), BJZZ gets all of the inferred trade sides incorrect. Finally, for 388,190 (almost 55%) of BJZZ-inferred retail orders with multiple fills, BJZZ infers different sides for different fills and produces an overall correct side 48.19% of the trades.

In Panel E of Table 9, we consider what happens when we restrict the sample to instances where the quoted spread is a penny. This is important because at spreads wider than a penny, the BJZZ approach to inferring order side is more error prone. Consider a stock with an NBB of $10.00 and an NBO of $10.01. A trade at $10.002 has a sub-penny increment of 0.2 and is (most likely) properly typed as a sell by BJZZ. Suppose that the NBO increases to $10.02 and a trade occurs at $10.012. Again, the sub-penny increment is 0.2 but it seems more likely that the trade is a buy. Without restricting the NBBO to a penny, the algorithm misclassifies the order side for 5.98% of the matched sample of retail trades. However, as shown in Panel E of Table 7, when we restrict our sample to matched retail trades received by the wholesaler(s) when the width of the NBBO is $0.01, the percentage of trades for which the inferred order side is incorrect falls to 0.60%.

To summarize, we find that several of the assumptions made to derive the BJZZ algorithm are suspect. At least during our sample period, there is a substantial portion of the retail trades potentially executed on an exchange by the data providing wholesaler(s). One can imagine that the sample of retail orders sent to exchanges to be executed is different than the sample of retail orders that are executed in the dark (e.g., away from exchanges). This, coupled with the fact that a large percentage of our sample of retail trades are executed either with no sub-penny price increment or with a sub-penny increment in the interval [0.4, 0.6] means that the BJZZ algorithm only has a chance of identifying a fraction of the initial sample of retail trades. In our case, less than one-third

of the presumed retail trades are identified as retail. Thus, BJZZ identifies a smaller fraction of known retail trades as retail than the fraction of some of our institutional trade samples it potentially identifies as retail. For the minority of known retail trades typed as retail by BJZZ, most are assigned the correct order side by the algorithm – although the algorithm struggles with multiple-trade orders.

V. Implications.

A. Order Imbalances.

The analysis thus far indicates that the BJZZ methodology fails to identify as retail trades nearly 70% of the actual retail trades obtained from our data provider(s) and frequently identifies known institutional trades as retail (in our admittedly limited institutional order/trade samples). The next step in the inquiry is to address whether these failures result in any substantive differences in important inferences. One issue of importance is the measure of order buy-sell imbalances as BJZZ go on to associate order imbalance measures constructed using their inferred retail trades with future stock price movements in an effort to judge the informativeness of retail order flow.[18] On page 2262 of their paper, BJZZ find that their inferred order imbalances and the order imbalances derived from a dataset containing Nasdaq-identified retail trades and known order side in 117 stocks executed on the Nasdaq stock exchange is 0.70.

In order to assess whether this relatively high correlation persists in our data, we compute each of the four order imbalance measures using the BJZZ-inferred retail trades and their inferred trade side and the data provider(s) do the same using the entirety of their retail order data (regardless of execution venue) with the known order side on a daily basis over a two-year period

[18] In a different approach to identifying potential retail trading activity, Bartlett et al (2022) estimates the number of reported one-share trades that result from Robinhood's and Drivewealth's fractional share trading programs arguing that these represent an alternative measure of retail trading. They conclude that liquidity and volatility are associated with lagged measures of retail trading activity.

of time. These calculations are done for all of the symbols traded by the data provider. $MROIBVOL_{i,t}$ is the signed difference between the retail buy volume and the retail sell volume normalized by the sum of the retail buy and sell volume for stock i on day t. $MROIBTRD_{i,t}$ is the signed difference between the number of retail buy trades and the number of retail sell trades normalized by the sum of retail buy and sell trades for stock i on day t. $ODDMROIBVOL_{i,t}$ and $ODDMROIBTRD_{i,t}$ are the imbalance measures computed using only odd lot trades in stock i on day t.

We then compute correlations between order imbalance measures on a stock-day basis and average these daily correlations over the 104-week sample period of August 3, 2020 to July 26, 2022. For any given stock-week observation, we eliminate all that are missing any of the eight (four based on the entire set of wholesaler(s) retail orders and four based on BJZZ's methodology of identifying retail trades in the same set of symbols traded by the wholesaler(s) in TAQ) or stock-weeks with any extreme order imbalance values (0, +1, or -1) as all of these indicate that there is a paucity of observations when computing the order imbalance measures. We begin with 1,021,091 observations. After eliminating stock-weeks where there is insufficient trading to obtain a reliable imbalance metric, we retain 823,621 observations. In Table 10, we report our experience with each of their order imbalance measures.

[Insert Table 10 about here.]

In Panel A, we see that our overall correlations between the BJZZ-inferred order imbalance and the comprehensive wholesaler(s) retail order imbalance is less than one-half of what BJZZ report. Volume-based measures are more highly correlated than trade-based metrics, but none exceed 0.35 correlation. Furthermore, we find that there is a bias in the BJZZ measures – they tend

to over-estimate the buy imbalance for the all-trade measures and under-estimate the buy imbalance for odd lot trade metrics.

In Panel B, we examine some descriptive statistics regarding the order imbalance (OIB) measures for the wholesaler(s) sample of all trades and the BJZZ-inferred measure. In column two (three), we report the mean (median) difference (wholesaler(s) minus BJZZ-inferred) in order imbalance measures with stock-week as the unit of observation. Consistent with Panel A, we find that the BJZZ-inferred measure is, on average, more (less) buy imbalance oriented than the actual wholesaler(s) measure for the imbalance measures focusing on total (odd lot) volume and trades. In columns four and five, we count the number of sample weeks for which the BJZZ measure is less than the measure computed using all wholesaler(s) retail trades in our sample and conduct a binomial test that the true proportion equals one-half (57 weeks). For the total volume- and trade-based measures, there is a clear bias of the BJZZ-inferred measure to be more buy oriented than the wholesaler(s) measure. For the odd lot volume-based measure there is the opposite bias but for the odd lot trade-based measure there appears to be no bias in a binomial test.

B. Association of lagged Order Imbalance Measures with Future Stock Returns.

Although the correlations between BJZZ inferred order imbalance measures and order imbalance measures computed with all of our proprietary data are not as high as BJZZ found when comparing their measures to a small sample of Nasdaq data, it is possible that we might still conclude that retail order imbalance measures are positively correlated with future close-to-close stock returns. To address this question, we recreate BJZZ's regression of week *t* order imbalance measures on week *t+1* returns as well as the control variables BJZZ suggest. That is, we estimate the following regression equation

$Return_{i,t} = \alpha + \beta_1\, Imbalance_{i,t-1} + \beta_2\, Return_{i,t-1} + \beta_3\, Return_{i,m-1} + \beta_4\, Return_{i,6m-1} +$

β_5 *Log Turnover*$_{i,m-1}$ + β_6 *Log Volatility*$_{i,m-1}$ + β_7 *Log Size*$_{i,m-1}$ + β_8 *Log B/M*$_{i,m-1}$ + ε,

where *Return*$_{i,t}$ equals stock *i*'s CRSP cumulative daily return for one of the 104 trading-week periods between August 3, 2020 and July 26, 2022, *Imbalance*$_{i,t-1}$ equals one of the four imbalance measures computed using either the BJZZ-inferred retail trades or the actual wholesaler(s)' trades for stock *i* from the week preceding the week used to compute *Return*$_{i,t}$, *Return*$_{i,t-1}$ is stock *i's* CRSP cumulative daily return for the trading week prior to the week used to compute *Return*$_{i,t}$, *Return*$_{i,m-1}$ is stock *i*'s CRSP cumulative daily return for the calendar month prior to the week used to compute *Return*$_{i,t}$, *Return*$_{i,6m-1}$ is stock *i*'s CRSP cumulative daily return for the calendar six-month period prior to the week used to compute *Return*$_{i,t}$, *Log Turnover*$_{i,m-1}$ is the log of the stock's monthly turnover in the calendar month prior to the week used to compute *Return*$_{i,t}$, *Log Volatility*$_{i,m-1}$ is the log of the stock's daily return volatility in the calendar month prior to the week used to compute *Return*$_{i,t}$, *Log Size*$_{i,m-1}$ is the log of the stock's market capitalization at the end of the calendar month prior to the week used to compute *Return*$_{i,t}$, and *Log B/M*$_{i,m-1}$ is the log of the book-to-market ratio at the end of the calendar month prior to the week used to compute *Return*$_{i,t}$.[19]

We run the regressions twice; first using all of the securities traded by the data provider(s) and, second, consistent with BJZZ, requiring the security traded be a common stock (CRSP code 10 or 11) and listed one of three exchanges (NYSE, NYSE MKT, and Nasdaq). Both sets of regressions are run over the 104-week (103 weeks considering the lagging of the OIB measures) sample period. We run Fama-Macbeth regressions with Newey West standard errors. The first set of regressions use 487,939 security-weeks of data and the BJZZ restrictions produce a sample size

[19] We follow Fama-French in the construction of book value of equity. When the book value of stockholder equity is available on the most recent Compustat annual file prior to Ret(i,t), that variable is used along with Compustat's year-end stock price and shares outstanding. If the book value of stockholder equity is unavailable on the most recent Compustat annual file prior to Ret(i,t), then we use total assets minus total liabilities from the most recent quarterly Compustat file prior to Ret(i,t). Should the most recent Compustat quarterly file be missing either of the necessary data items, we try successive older Compustat quarterly files up to one year prior to Ret(i,t) before deleting the observation from the regression.

of 304,953 stock-weeks. We report the results of the regression in Table 11 for Newey West lag 5 as our results are robust to alternative numbers of lags. Panel A reports results for the inclusive sample securities and Panel B the restricted sample consistent with BJZZ.

[Insert Table 11 about here.]

In Panel A of Table 11, the estimated coefficients on the prior week's OIB measure is reliably greater than zero for both the wholesaler and BJZZ-inferred versions of the OIB metrics only for *MROIBVOL*. For the *MROIBTRD* measure, only the imbalance measure constructed from the wholesaler(s) data is positively associated with future returns. Neither odd-lot OIB metric is significantly associated with future return regardless of how it is constructed. Thus, for our baseline results, because the conclusions are generally identical between the wholesaler and BJZZ OIB metrics, we conclude that BJZZ Type I and Type II errors appear to minimally affect the inferred association between returns and lagged OIB measures. Although we conclude a positive association regardless of the potential for errors in the BJZZ methodology for the volume-based, all-trade metric, we reach conflicting conclusions using the wholesaler data than we find using the BJZZ methodology for the trade-based all-trade measure. Finally, our results on the control variables parallel very closely those reported by BJZZ in their Table III.

In Panel B of Table 11, we report the results of running the regressions after restricting our sample securities to common stocks listed on the NYSE, NYSE MKT or Nasdaq – as in BJZZ. This reduces the sample size by roughly 37.5% and produces results more consistent with but somewhat weaker than those reported by BJZZ. As with the all-trade OIB measures, the wholesaler(s) OIB metric is statistically significantly associated with future returns using both the trade-based and the volume-based all-trade measures. The BJZZ-inferred OIB metric is strongly significant only with the volume-based all-trade measure but marginally significant for the trade-

based measure as well. In the restricted sample, we find that both of the wholesaler(s) odd-lot OIB metrics are positively correlated with future stock returns but, again, only the volume-based odd-lot measure using BJZZ. As in BJZZ, the intercept terms are statistically significant after restricting the sample securities to common stocks listed on one of the three exchanges. Thus, the Type I and Type II errors inherent in BJZZ appear to be more of a problem with the trade-based versus volume-based OIB metrics. Again, the control variable regression results are similar to BJZZ.

Baradehi et al. (2022) conclude that institutional order flow drives the statistical association between lagged retail order imbalances and stock returns because most wholesalers also are electronic liquidity providers on many execution venues (e.g., exchanges, ATSs, and single-dealer platforms) on which they interact with institutional orders. As institutions demand one-sided liquidity, wholesalers internalize opposite-sided retail orders in an attempt to manage inventory risk. As the market unwinds the price pressure caused by the institutional orders over the intervening week, these offsetting retail orders are identified as informed by BJZZ. Baradehi et al. (2022) posit that if wholesalers' order imbalance measures included both internalized and externalized order flow, that would add noise to the statistical association with returns and diminish the significance found by BJZZ. Our data of wholesaler(s) OIB measures include both internalized and externalized orders when constructing their order imbalances, yet we find that the wholesaler(s)' measures continue to exhibit a positive correlation to weekly returns with the volume-based measures. This finding conflicts with the Baradehi et al. (2022) hypothesis.

In summary, although the correlation between the order imbalance measures using the wholesaler retail orders and the order imbalance measures using the BJZZ-inferred retail orders is less than one-half that reported by BJZZ between their inferred retail trades and a small sample of known Nasdaq retail trades, the inferences regarding the apparent informativeness of retail order

flow, when statistically significant, generally is the same. In fact, the wholesaler OIB metrics are more robustly related to future stock returns than the BJZZ-inferred measures. This latter result conflicts with the statement in Baradehi et al (2022) that including externalized wholesaler trades would dilute/eliminate the correlation with future returns as it would add noise to the inferred institutional order imbalance.

C. Regressions based on Retail Trading Intensity – Controlling for Type I Errors.

Finally, we recompute the regression results with an attempt to control for Type I errors (i.e., identifying institutional trades as retail trades). To do so, we use SEC Rule 605 reports to rank stocks by the fraction of marketable order shares executed by the six major wholesalers relative to TAQ consolidated share trading volume.[20] Using the retail portion of share volume in the prior month for securities with sufficient (10,000 shares in the month) volume, we form quintiles and re-estimate the BJZZ regression using only stocks from a given retail-intensity quintile. Specifically, we take the following steps: 1.) require that the symbol have at least 10,000 shares executed in month t and a valid fractional-retail measure, 2.) if desired, impose the BJZZ restrictions for only common stocks listed on BJZZ-selected exchanges, and 3.) require that at least one valid order imbalance measure exist in month t+1. At that point, we rank the retail intensity measures, assign month t quintiles, merge the weekly panel data of weekly returns and BJZZ control variables and perform Fama-Macbeth regressions. We posit that the highest quintile of retail trading fraction has less chance of a Type I error as it has a smaller fraction of non-retail trading than does quintile 1. On average, the quintiles contain 1,604 symbols without the BJZZ constraints and 633 stocks with the constraints.

[Insert Table 12 about here.]

[20] The six wholesalers are Citadel Securities, G1X Susquehanna, Jane Street, Two Sigma, UBS, and Virtu Financial.

We estimate the regressions by retail-intensity quintile for the four BJZZ order imbalanced measures (again, once using BJZZ-inferred retail trades and once using the data-providing wholesaler(s) retail trades) and report the results in Table 12. In Panel A of Table 12 we report the results for the sample using all symbols traded by the data providing wholesaler(s) and, in Panel B, we report the results for the BJZZ-consistent sample of common stocks listed on one of the three exchanges. To conserve space, we report only the coefficient estimates and standard errors for the OIB metrics.

We first look at Panel A regression results, which include all securities traded by retail investors. Examining the all-trade and all-volume OIB measures in columns (1) – (4), we find a large disagreement between the BJZZ-inferred measures and the wholesaler(s) results. For the first four quintiles, the wholesaler(s) lagged order imbalance measures are statistically associated with returns in quintiles 2 and 3 but not quintile 1 while the BJZZ-inferred measures are insignificant in quintiles 1-3. Both are significant in quintile 4. For the highest retail intensity quintile both BJZZ-inferred and (weakly) the wholesaler(s)' OIB metrics are significantly positive. Turing to the odd-lot OIB measures, the BJZZ and wholesaler(s) regression results provide consistent conclusions – no reliable relation between lagged OIB measures and returns for retail-intensity quintiles 1 thorough 4 but (at least weakly) statistical association for the most heavily retail-traded quintile.

Turning to Panel B, which focuses on the subset of securities examined in BJZZ, we find that both the BJZZ-inferred and the wholesaler(s) lagged OIB measures are at least weakly significantly positively correlated with returns for the highest-retail intensity quintile. For the all-trade and all-volume OIB metrics, the wholesaler(s)' lagged metrics are significantly positively

correlated with returns for all but the lowest retail intensity quintile but the BJZZ measures exhibit no statistically reliable association for any of the lower quintile portfolios.

We know that the wholesaler order imbalance metrics have no Type II errors and we suspect only minimal Type I errors (some institutions use retail brokers to avoid commissions – see Battalio and Jennings (2022)). The BJZZ order imbalance metrics have both types of errors but there should be relatively fewer Type I errors in retail intensity quintile 5. The fact that the all-trade OIB results generally differ between BJZZ and wholesaler OIBs in lower retail intensity quintiles but agree in the highest quintile suggests that Type I errors might be important in the BJZZ methodology for these OIB metrics. It is possible that using the odd-lot OIB measures help mitigate the effect of Type I errors. The BJZZ and wholesaler OIB measures exhibit considerably more agreement (especially in the larger, more retail representative security sample).[21]

V. Conclusion.

Given the wide-ranging interest in the activities of retail investors in the academic literature and the professional and regulatory spheres, it is important that researchers properly identify retail trades in publicly available data in order to draw correct conclusions for policy formation.[22] Boehmer et al (2021) provide a methodology to infer retail trades from TAQ data that has been used in numerous recent academic and practitioner studies. We undertake a study of the accuracy of their assumptions in designing the methodology and conduct one analysis of the implications of documented inaccuracies. Using proprietary data, we demonstrate that the BJZZ methodology correctly identifies as retail less than one-third of trades that are commonly thought of as retail.

[21] In an attempt to see if Type I errors matter, we reran the base regressions including: 1.) all odd lot trades and 2.) all trades with execution prices at the full penny but within the NBBO (i.e., all full penny trades that are price improved). Revised results do not differ in a meaningful way from the base results. We conclude that these attempts to (potentially) include more retail orders in the regressions are unsuccessful.

[22] Battalio and Jennings (2023) argue that the use of BJZZ-identified retail trades rather than a sample of all retail trades results in lower estimates of the potential costs of failed auctions in the SEC's proposed Order Competition Rule.

Furthermore, we document that it is frequently the case that known institutional trades that BJZZ assume will not be identified as retail by their methodology are indeed inferred to be retail trades. Do these Type I and Type II errors lead to important incorrect inferences? We follow BJZZ to examine one such research path. Specifically, we compare the order imbalance measures computed with a known set of retail orders and compare that to the order imbalance measures computed using BJZZ methodology to infer retail trades. The order imbalance measures demonstrate a relatively low correlation with each other – less than one-half the correlation that BJZZ find with a limited sample of Nasdaq proprietary data. When taking the two sets of order imbalance metrics to the data to investigate their association with security returns, we find reasonable agreement when examining on the subset of securities that were the focus of the BJZZ analysis (common stocks listed on select exchanges). In an effort to create an environment where we might be able to assess the importance of Type I errors (false positives with BJZZ) and Type II errors (false negatives with BJZZ), we create quintiles of the sample securities based on an estimate of the fraction of total trading that is from retail investors. In these subsamples, the more retail intense the trading, the less importance Type I errors play – the bulk of the trading is indeed retail. Here we find evidence consistent with Type I error being important for the all-trade order imbalance measures. In these cases, the wholesaler lagged order imbalance measure is generally significantly associated with returns but the BJZZ measures are significant only for the highest retail intensity quintile.

References

Baradehi, Yashar, Dan Bernhardt, Zhi Da, and Mitch Warachka, 2022. Uncovering the liquidity premium in stock returns using retail liquidity provision, SSRN paper number 4057713.

Barber, Brad, Xing Huang, Phillipe Jorion, Terrance Odean and Christopher Schwartz, 2022. A (sub)penny for your thoughts: Tracking retail investor activity in TAQ, SSRN paper number 4202874.

Barber, Brad, Xing Huang, Terrance Odean, and Christopher Schwarz, 2021. Attention induced trading and returns: Evidence from Robinhood users, forthcoming in the *Journal of Finance*.

Bartlett, 2022, Modernizing odd lot trading, *Columbia Business Law Review*, forthcoming.

Bartlett, Robert, J. McCrary, and Maureen O'Hara, 2022, Tiny trades, bug questions: Fractional shares, SSRN paper number 4248449.

Battalio, Robert, and Robert Jennings, 2022. Why do brokers who do not charge payment for order flow route marketable to Wholesalers, SSRN paper number 4304124.

Battalio, Robert, and Robert Jennings, 2023. On the potential costs of mandating qualified auctions for marketable retail orders, Mendoza College of Business working paper.

Battalio, Robert, Brian Hatch, and Mehmet Saglam, 2022. The cost of exposing large institutional orders to electronic liquidity providers, forthcoming in *Management Science*.

Blankespoor, Elizabeth, Ed deHaan, and Christina Zhu, 2018. Capital market effects of media synthesis and dissemination: Evidence from robo-journalism, *Review of Accounting Studies* 23, 1 – 36.

Boehmer, Ekkehart, Charles Jones, Xiaoyan Zhang, and Xinran Zhang, 2021. Tracking retail investor activity, *Journal of Finance* 2249 – 2305.

Bonsall IV, S. B., Green, J., & Muller III, K. A. (2020). Market uncertainty and the importance of media coverage at earnings announcements. *Journal of Accounting and Economics*, *69*(1), 101264.

Bushee, B., Cedergren, M., & Michels, J. (2020). Does the media help or hurt retail investors during the IPO quiet period? *Journal of Accounting and Economics*, *69*(1), 101261.

Bradley, Daniel, Russell James, and Jared Williams, 2022. Non-deal roadshows, informed trading, and analyst conflicts of interest, *Journal of Finance* 77, 265 – 315.

Cready, William, Abdullah Kumas, and Musa Subasi, 2014. Are trade size-based inferences about traders reliable? Evidence from institutional earnings-related trading, *Journal of Accounting Research* 52, 877 – 909.

Farrell, M., Green, T. C., Jame, R., & Markov, S. (2022). The democratization of investment research and the informativeness of retail investor trading. *Journal of Financial Economics*, *145*(2), 616-641.

Guest, N. M. (2021). The information role of the media in earnings news. *Journal of Accounting Research*, *59*(3), 1021-1076.

Israeli, D., Kasznik, R., & Sridharan, S. A. (2022). Unexpected distractions and investor attention to corporate announcements. *Review of Accounting Studies*, *27*(2), 477-518.

Hendershott, T., D. Livdan, and N. Schurhoff, 2015, Are institutions informed about news? *Journal of Financial Economics* 117, 249-287.

Hvidkjaer, Soeren, 2008. Small trades and the cross-section of stock returns, *Review of Financial Studies* 21, 1123 – 1151.

Kaniel, Ron, Gideon Saar, and Sheridan Titman, 2008. Individual investor trading and stock returns, *Journal of Finance*, 273 – 310.

Kelley, Eric and Paul Tetlock, 2013. How wise are crowds? Insights from retail orders and stock returns, *Journal of Finance* 68, 1229 – 1265.

Lee, Charles, 1992. Earnings news and small traders, *Journal of Accounting and Economics* 15, 265 – 302.

Lee, Charles and Mark Ready, 1991. Inferring trade direction from intraday data, *Journal of Finance* 46, 733 – 746.

Lee, Charles and Balkrishna Radhakrishna. 2000. "Inferring Investor Behavior: Evidence from TORQ Data." *Journal of Financial Markets* 3, 83–111.

Odean, Terrance, 1998. Are investors reluctant to realize their losses?, *Journal of Finance,* 1775 - 1798.

Reilly, Frank, 1979. <u>Investment Analysis & Portfolio Management,</u> The Dryden Press, Hinsdale, Illinois.

Trading Talk – Market Structure Analysis: An Update on Retail Market Share in US Equities, June 24, 2022.

Table 1. Non-retail Order Flow Trading on Sub-penny Prices for Sample Stocks in December 2021.

The data provider(s) identified 2,100,769 non-retail, sub-penny but not half-penny executions. Requiring that the trades occur in regular market hours reduced the sample to 2,100,162 observations.

Panel A. Descriptive Statistics.

Variable	Mean	Median	Minimum	Maximum
Execution Time	12:21:42	11:55:49	9:30:00	16:00:00
Shares Filled	302	100	1	1,600,000
Trade Price	$112.28	$38.61	$1.0005	$5850.0005

Panel B. Percent of Trades by Time of Trading Day.

Time Interval	Percent of Trades
9:30:00 to < 10:00:00	17.85%
10:00:00 to < 11:00:00	19.54%
11:00:00 to < 12:00:00	13.43%
12:00:00 to < 13:00:00	10.24%
13:00:00 to < 14:00:00	9.42%
14:00:00 to < 15:00:00	10.50%
15:00:00 to 16:00:0	19.01%

Panel C. Percent of Trades by Trade Size.

Trade Size Interval	Percent of Trades
1 – 99 shares	37.17%
100 – 499 shares	49.25%
500 – 999 shares	7.08%
1000 – 1999 shares	3.86%
2000 – 4999 shares	2.01%
5000 – 9999 shares	0.51%
> 9999 shares	0.14%

Table 1 (continued).

Panel D. Percent of Trades by Trade Price.

Trade Price Interval	Percent of Trades
≥ \$1 and < \$10.00	20.84%
≥ \$10 and < \$50	35.29%
≥ \$50 and < \$100	13.73%
≥ \$100 and < \$250	18.62%
≥ \$250 and < \$500	8.88%
\$500+	2.65%

Table 2. Distribution of Sub-penny Increments for a Sample of 2,100,162 Non-Retail Trades in December 2021 by Trade Side.

Sub-penny increments are defined by BJZZ as Z = 100*mod(price,.01).

Interval	Buy Orders	Sell Orders
$0 < Z \leq .1$	39.43%	2.66%
$.1 < Z \leq .2$	21.75%	3.73%
$.2 < Z \leq .3$	6.86%	3.02%
$.3 < Z \leq .4$	9.14%	9.09%
$.4 < Z \leq .5$	1.61%	1.03%
$.5 < Z \leq .6$	6.46%	6.60%
$.6 < Z \leq .7$	4.93%	4.42%
$.7 < Z \leq .8$	7.02%	15.48%
$.8 < Z \leq .9$	2.34%	20.43%
$.9 < Z < 1.00$	0.46%	33.54%

Table 3. Sub-penny pricing distribution for a sample of institutional trades executed away from exchanges between January 2011 and March 2012.

We obtain a sample of trades generated by institutional orders executed by a large investment bank (IB) between January 2011 and March 2012 used by Battalio et al. (2022). From this sample of trades, we extract all trades executed by an electronic liquidity provider (ELP) like Getco, Citadel, and Knight Securities and all trades executed in the IB's dark pool.

	% of Trades	
100*mod(Price,0.01)	Trades Executed in IB's Dark Pool (N = 166,266)	Trades Executed by ELP (N = 136,822)
0.00	38.4%	21.2%
0.00 < mod < 0.40	17.1%	38.7%
0.40 ≤ mod ≤ 0.60	28.4%	0.5%
0.60 < mod < 1.00	16.1%	39.6%

Table 4. Sub-Penny Pricing Increments Associated with the Public Employees' Retirement Association of Colorado in 2016 and 2017.

This table provides a breakdown of sub-penny pricing increments into the percentage done at the whole penny level, percentage done at what BJZZ infer as the half-penny and all other sub-penny prices.

Panel A. All Facilitating Broker's Own Dark Pool Trades (70,739 trades).

Fraction of the Penny	% of Trades
0.000000	35.8%
0.000001 to 0.003999	7.2%
0.004000 to 0.006000	48.8%
0.006001 to 0.009999	8.4%

Panel B. All Other Dark Pool Trades (292,720 trades).

Fraction of the Penny	% of Trades
0.000000	54.3%
0.000001 to 0.003999	1.9%
0.004000 to 0.006000	42.0%
0.006001 to 0.009999	1.9%

Panel C. All Electronic Liquidity Provider Trades (6,203 trades).

Fraction of the Penny	% of Trades
0.000000	50.4%
0.000001 to 0.003999	9.0%
0.004000 to 0.006000	24.1%
0.006001 to 0.009999	17.4%

Table 5. Descriptive statistics for the 2,741 sample securities.

From the 9,584 unique security symbols traded by our data provider(s) for the month of December 2020, we select symbols that average at least 100 trades per trading day and have no days without trades. Finally, to reduce the frequency with which securities can be quoted in sub-penny increments, we require a closing security price of greater than $1.00 at month's end. Statistics in the table are equally-weighted across sample stocks.

Statistic	End-of-Month Price	Round-Lot Trades in Month
Mean	$64.57	23,328
Median	$25.86	6,274
Minimum	$1.01	2,201
Maximum	$3,255.63	1,967,137

Table 6. Descriptive statistics for the retail trades in our proprietary dataset reported to FINRA's Trade Reporting Facility (TRF) and for those retail trades we could match to TAQ TRF trades for the ten trading days from December 1 through December 14, 2020.

Dark Retail Trades are retail trades from our proprietary data reported to FINRA's TRF. We match to TAQ TRF trade data based on security symbol, trade price and size, and trade time. Symbol, trade size, and trade price require exact matches. We allow ten milliseconds difference between the TAQ Participant Timestamp and the data provider(s)' timestamp. The average time difference between the TAQ participant time of the matched TAQ trade and the execution time of the retail trade is 0.00121 seconds. We are able to match 78.16% of our proprietary data retail trades to TRF trades in the TAQ database.

Panel A. Descriptive statistics for the original and matched samples of retail trades.

Variable	Mean	
	Dark Retail Trades (N = 30,881,022)	Matched Retail Trades (N = 24,135,132)
Execution Time	12:13:50	12:15:00
Trade Quantity	223 shares	231 shares
Trade Price (trade-weighted)	$113.19	$123.94
Mean Share Price (symbol-weighted)	$62.49	$62.47
Order Quantity	2,307 shares	784 shares
% Buy Orders	55.84%	56.82%
% Short Sell Orders	3.17%	2.84%

Panel B. Distribution of trade times throughout the trading day for the original and matched sample of retail trades. There are 812 trades occurring at exactly 16:00:00 in the proprietary data that we do not match to TAQ.

Hour	Dark Retail Trades (N = 30,881,022)	Matched Retail Trades (N = 24,135,132)
9:30 to 10:00	15.96%	15.57%
10:00 to 11:00	21.46%	21.36%
11:00 to 12:00	14.34%	14.39%
12:00 to 1:00	11.95%	12.07%
1:00 to 2:00	11.43%	11.53%
2:00 to 3:00	10.86%	10.98%
3:00 to 4:00	14.00%	14.11%

Table 7. Sub-penny pricing distribution for matched retail trades by trade side conditional on the order receipt time National Best Bid and Offer spread's width.

We match proprietary wholesaler(s)'s trades to TAQ TRF trade data based on security symbol, trade price and size, and trade time. Symbol, trade size, and trade price require exact matches. We allow ten milliseconds difference between the TAQ Participant Timestamp and the data provider(s)' timestamp.

	All Spreads		NBBO Width = $0.01	
100*mod(Price,0.01)	Buy Orders (N=13,714,104)	Sell Orders (N=10,421,028)	Buy Orders (N=5,977,765)	Sell Orders (N=4,556,201)
0.00	21.42%	18.24%	13.06%	12.82%
0.00<mod≤0.10	0.34%	9.98%	0.15%	11.33%
0.10<mod≤0.20	0.29%	0.99%	0.00%	1.46%
0.20<mod≤0.30	0.33%	0.63%	0.03%	0.79%
0.30<mod<0.40	0.19%	0.63%	0.00%	0.73%
0.40≤mod≤0.60	17.60%	12.30%	23.80%	16.09%
0.60<mod≤0.70	0.83%	0.20%	0.97%	0.00%
0.70<mod≤0.80	0.95%	0.21%	1.05%	0.02%
0.80<mod≤0.90	1.52%	0.22%	2.01%	0.00%
0.90<mod<1.00	13.34%	0.11%	15.68%	0.01%
0.00<mod<0.40	1.15%	12.23%	0.18%	14.31%
0.60<mod<1.00	16.64%	0.74%	19.71%	0.01%

Table 8. Descriptive statistics comparing BJZZ-identified retail trades to the sample of retail trades matched to TAQ trades.

BJZZ require that the trade price be on one of two sub-penny intervals; greater than zero and less than $0.004 or greater the $0.006 and less than $1.000. In addition, BJZZ require that the trade price exceed $1.00 and not be subject to any non-normal trade condition. The BJZZ methodology results in identifying 7,349,520 of 24,135,132 (30.45%) of TAQ-matched proprietary data known retail trades.

Panel A. Descriptive Statistics of BJZZ-Identified retail trades and all matched retail trades.

Variable	Mean	
	Matched Retail Trades (N = 24,135,132)	BJZZ-Identified Retail Trades (N = 7,349,520)
Execution Time	12:15:00	12:19:27
Execution Size	231 shares	296 shares
Trade Price (trade-weighted)	$123.14	$108.88
Trade Price (stock-weighted)	$62.47	$63.18
Order Quantity	784 shares	582 shares
Order Side – Percent Buys	56.87%	57.96%
Percent Short Sells	2.84%	3.94%

Panel B. Distribution of trade times throughout the trading day for the BJZZ-Identified retail trades and all matched retail trades.

Hour	Matched Retail Trades (N = 24,135,132)	BJZZ-Identified Retail Trades (N = 7,349,520)
9:30 to 10:00	15.61%	12.72%
10:00 to 11:00	21.15%	23.68%
11:00 to 12:00	14.43%	14.37%
12:00 to 1:00	12.10%	11.59%
1:00 to 2:00	11.56%	11.04%
2:00 to 3:00	11.01%	10.75%
3:00 to 4:00	14.15%	15.85%

Panel C. BJZZ-Identified retail trades a percentage of the matched retail trades by time of day.

Hour	BJZZ Success Rate
9:30 to 10:00	24.98%
10:00 to 11:00	33.75%
11:00 to 12:00	30.41%
12:00 to 1:00	29.25%
1:00 to 2:00	29.16%
2:00 to 3:00	29.83%
3:00 to 4:00	34.20%

Panel D. BJZZ-Identified retail trades as a percentage of matched retail trades by order size.

Order Size (shares)	BJZZ Success Rate
1 – 99	28.13%
100 – 499	32.49%
500 – 999	33.31%
1000 – 1999	34.56%
2000 – 4999	45.74%
> 4999	26.52%

Panel E. BJZZ-Identified retail trades as a percentage of matched retail trades by trade size.

Trade Size (shares)	BJZZ Success Rate
1 – 99	28.45%
100 – 499	32.88%
500 – 999	32.17%
1000 – 1999	32.64%
2000 – 4999	48.74%
> 4999	48.34%

Panel F. BJZZ-Identified retail trades as a percentage of matched retail trades by trade price.

Execution Price	BJZZ Success Rate
$1.00 – $9.9999	33.93%
$10.00 – $49.9999	30.70%
$50.00 – $99.9999	39.95%
$100.00 – $249.9999	28.75%
$250.00 – $499.9999	28.89%
$500.00+	25.59%

Table 9. Descriptive statistics regarding the success of BJZZ trade side inference.

BJZZ infer trade side from the sub-penny pricing increment. Those trades with sub-penny pricing greater than $0.000 and less than $0.004 are inferred sells and those trades with sub-penny pricing greater than $0.006 and less than $0.01 are inferred buys. We compare the BJZZ-inferred side to the known order side in the proprietary retail data.

Panel A. Overall mix of the actual trade side to BJZZ inferred trade side.

	Known Order Side from Matched Retail Trade	
BJZZ Inferred Side	Buy	Sell
Buy	54.37%	2.29%
Sell	3.59%	39.74%

Panel B. Percentage of matched retail trades with the correct BJZZ-inferred trade side by time of trading day.

Hour	Correct Trade Side
9:30 to < 10:00	93.77%
10:00 to < 11:00	95.50%
11:00 to < 12:00	94.54%
12:00 to < 13:00	94.12%
13:00 to < 14:00	93.96%
14:00 to < 15:00	93.99%
15:00 to 16:00	95.11%

Panel C. Percentage of matched retail trades with the correct BJZZ-inferred trade side by order size and trade size.

Size (shares)	Order Size	Trade Size
1 – 99	94.24%	93.71%
100 – 499	93.72%	94.56%
500 – 999	94.09%	97.42%
1000 – 1999	94.56%	98.56%
2000 – 4999	99.13%	99.61%
> 4999	99.28%	98.82%

Panel D. Percentage of matched retail trades with the correct BJZZ-inferred trade side for matched retail orders filled with one and with multiple executions.

	Number of Orders	Success Rate
Retail orders filled with one trade	5,971,948	94.85%
Retail orders filled with multiple trades	711,641	62.62%

Panel E. Overall mix of the actual trade side to BJZZ inferred trade side when order placement time quote is $0.01.

	Known Order Side from Matched Retail Trade	
BJZZ Inferred Side	Buy	Sell
Buy	57.56%	0.07%
Sell	0.53%	41.84%

Table 10. Order imbalance descriptive statistics.

$MROIBVOL_{i,t}$ is the signed difference between the retail buy volume and the retail sell volume normalized by the sum of the retail buy and sell volume for stock *i* on day *t*. $MROIBTRD_{i,t}$ is the signed difference between the number of retail buy trades and the number of retail sell trades normalized by the sum of retail buy and sell trades for stock *i* on day *t*. $ODDMROIBVOL_{i,t}$ is the signed difference between the retail odd lot buy volume and the retail odd lot sell volume normalized by the sum of the retail buy and sell volume for odd lot trades in stock *i* on day *t*. $ODDMROIBTRD_{i,t}$ is the signed difference between the number of retail odd lot buy trades and the number of retail odd lot sell trades normalized by the sum of retail buy and sell trades for odd lot trades in stock *i* on day *t*. We construct each measure separately for the entire sample of proprietary retail trades and for the retail trades identified in the TAQ database by the BJZZ methodology. Our sample period is August 3, 2020 through July 26, 2022. We characterize the daily across stock correlations between the statistics computed using the inferred and the actual retail trading data in the table below.

Panel A. Overall Correlation Statistics

OIB Measure	Correlation	Mean Differences (Wholesaler(s) data – BJZZ*
MROIBVOL	.3415	-0.0129
MROIBTRD	.3021	-0.0557
ODDMROIBVOL	.3181	+0.0324
ODDMROIBTRC	.2937	+0.0327

* All differences are statistically significant at beyond the .0001 level with a standard t-test.

Panel B. Weekly Order Imbalance (OIB) Statistics

Measure	Weekly Across-Stock Wholesaler(s) – Inferred OIB		# Weeks with Mean Across-Stock Wholesaler(s) OIB > Mean Inferred OIB	# Weeks where Binomial Test Indicates Wholesaler(s) OIB > Inferred OIB Significance level in parenthesis (Hypothesis is ρ = .5)
	Mean	Median		
MROIBVOL	-0.0094	-0.0122	27 out of 104	< 0.0001
MROIBTRD	-0.0082	-0.0105	0 out of 104	< 0.0001
ODDMROIBVOL	0.0764	0.0766	102 out of 104	< 0.0001
ODDMROIBTRD	s.0767	0.0770	56 out of 104	.1634

Table 11 – BJZZ Regressions examining the association between lagged order imbalance measures and security returns.

We estimate the following regression equation, $Return_{i,t} = \alpha + \beta_1\ Imbalance_{i,t-1} + \beta_2\ Return_{i,t-1} + \beta_3\ Return_{i,m-1} + \beta_4\ Return_{i,6m-1} + \beta_5\ Log\ Turnover_{i,m-1} + \beta_6\ Log\ Volatility_{i,m-1} + \beta_7\ Log\ Size_{i,m-1} + \beta_8\ Log\ B/M_{i,m-1} + \varepsilon,$ where $Return_{i,t}$ equals stock i's CRSP cumulative daily return for one of the four trading-week periods, $Imbalance_{i,t-1}$ equals one of the four imbalance measures computed using BJZZ-inferred retail trades for stock i from the week preceding the week used to compute $Return_{i,t}$, $\mathrm{Return}_{i,t-1}$ is stock i's CRSP cumulative daily return for the trading week prior to the week used to compute $Return_{i,t}$, $Return_{i,m-1}$ is stock i's CRSP cumulative daily return for the calendar month prior to the week used to compute $Return_{i,t}$, $Return_{i,6m-1}$ is stock i's CRSP cumulative daily return for the calendar six-month period prior to the week used to compute $Return_{i,t}$, $Log\ Turnover_{i,m-1}$ is the log of the stock's monthly turnover in the calendar month prior to the week used to compute $Return_{i,t}$, $Log\ Volatility_{i,m-1}$ is the log of the stock's daily return volatility in the calendar month prior to the week used to compute $Return_{i,t}$, $Log\ Size_{i,m-1}$ is the log of the stock's market capitalization at the end of the calendar month prior to the week used to compute $Return_{i,t}$, and $Log\ B/M_{i,m-1}$ is the log of the book-to-market ratio at the end of the calendar month prior to the week used to compute $Return_{i,t}$. We compute order imbalance measures on a stock-day basis and, for the regressions presented in Panels A, B, and D, average those daily correlations on a weekly basis over the 104-week sample period of August 3, 2020 to July 26, 2022. For any given stock-week observation, we eliminate all that are missing any of the eight (four based on the wholesaler(s)'s retail orders and four based on BJZZ's methodology of identifying the same set of stocks in TAQ) or stock-weeks with any extreme order imbalance values (0, +1, or -1) as all of these indicate that there is a paucity of observations when computing the order imbalance measures. The regressions in Panels A and B have 823,621 stock-week observations. In Panel C, the unit of observation is a stock-day. We run Fama-Macbeth regressions with Newey West standard errors. We report the results of the regression in Table 11 for Newey West lag = 5 as additional lags do not change our conclusions. *, **, and *** indicates statistical significance at the 0.10, 0.05, and 0.01 level, respectively.

Panel A. Using all securities traded by the data provider.

	BJZZ	Actual	BJZZ	Actual	BJZZ	Actual	BJZZ	Actual
Intercept	0.0090	0.0100	0.0089	0.0100	0.0090	0.0094	0.0090	0.0094
MROIBTRD	0.0013 (1.25)	0.0027*** (4.90)						
MROIBVOL			0.0019*** (3.24)	0.0027*** (4.80)				
ODDMROIBTRD					0.0006 (0.45)	0.0017 (1.37)		
ODDMROIBVOL							0.0010 (1.11)	0.0017 (1.35)
$Return_{i,t-1}$	-0.0156***	-0.0177***	-0.0157***	-0.0178***	-0.0157***	-0.0172***	-0.0156***	-0.0172***
$Return_{i,m-1}$	0.0012	0.0008	0.0012	0.0008	0.0012	0.0006	0.0012	0.0006
$Return_{i,6m-1}$	0.0031	0.0028	0.0031	0.0028	0.0031	0.0028	0.0031	0.0028
Log $Turnover_{i,m-1}$	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Log $Volatility_{i,m-1}$	-0.0628	-0.0616	-0.0629	-0.0616	-0.0632	-0.0616	-0.0634	-0.0616
Log $Size_{i,m-1}$	-0.0003	-0.0004	-0.0003	-0.0004	-0.0003	-0.0003	-0.0003	-0.0003
Log $B/M_{i,m-1}$	0.0011	0.0012	0.0012	0.0012	0.0012	0.0012	0.0012	0.0012

Panel B. Using only common stocks (CRSP security codes 10 or 11) listed on multiple exchanges.

	BJZZ	Actual	BJZZ	Actual	BJZZ	Actual	BJZZ	Actual
Intercept	0.0145**	0.0149**	0.0144**	0.0149**	0.0145**	0.0147**	0.0145**	0.0147**
MROIBTRD	0.0022* (1.78)	0.0030*** (4.74)						
MROIBVOL			0.0016*** (2.63)	0.0029*** (4.49)				
ODDMROIBTRD					0.0016 (1.35)	0.0020** (2.52)		
ODDMROIBVOL							0.0018** (2.28)	0.0020** (2.49)
$Return_{i,t-1}$	-0.0143***	-0.0145***	-0.0144***	-0.0145***	-0.0142***	-0.0135***	-0.0143***	-0.0135***
$Return_{i,m-1}$	-0.0019	-0.00116	-0.0020	-0.0017	-0.0019	-0.0018	-0.0019	-0.0018
$Return_{i,6m-1}$	0.0017	0.0017	0.0017	0.0017	0.0017	0.0017	0.0017	0.0017
Log $Turnover_{i,m-1}$	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Log $Volatility_{i,m-1}$	-0.0684	-0.0692	-0.0683	-0.0691	-0.0684	-0.0698	-0.0687	-0.0698
Log $Size_{i,m-1}$	-0.0006	-0.0006	-0.0006	-0.0006	-0.0006	-0.0006	-0.0006	-0.0006
Log $B/M_{i,m-1}$	0.0015	0.0015	0.0015	0.0015	0.0015	0.0015	0.0015	0.0015

Table 12 – Estimating the stock return regressions using retail intensity quintile sub-samples.

We estimate the following regression equation, $Return_{i,t} = \alpha + \beta_1\ Imbalance_{i,t-1} + \beta_2\ Return_{i,t-1} + \beta_3\ Return_{i,m-1} + \beta_4\ Return_{i,6m-1} + \beta_5\ Log\ Turnover_{i,m-1} + \beta_6\ Log\ Volatility_{i,m-1} + \beta_7\ Log\ Size_{i,m-1} + \beta_8\ Log\ B/M_{i,m-1} + \varepsilon,$ where $Return_{i,t}$ equals stock i's CRSP cumulative daily return for one of the four trading-week periods, $Imbalance_{i,t-1}$ equals one of the four imbalance measures computed using BJZZ-inferred retail trades for stock i from the week preceding the week used to compute $Return_{i,t}$, $\mathrm{Return}_{i,t-1}$ is stock i's CRSP cumulative daily return for the trading week prior to the week used to compute $Return_{i,t}$, $Return_{i,m-1}$ is stock i's CRSP cumulative daily return for the calendar month prior to the week used to compute $Return_{i,t}$, $Return_{i,6m-1}$ is stock i's CRSP cumulative daily return for the calendar six-month period prior to the week used to compute $Return_{i,t}$, $Log\ Turnover_{i,m-1}$ is the log of the stock's monthly turnover in the calendar month prior to the week used to compute $Return_{i,t}$, $Log\ Volatility_{i,m-1}$ is the log of the stock's daily return volatility in the calendar month prior to the week used to compute $Return_{i,t}$, $Log\ Size_{i,m-1}$ is the log of the stock's market capitalization at the end of the calendar month prior to the week used to compute $Return_{i,t}$, and $Log\ B/M_{i,m-1}$ is the log of the book-to-market ratio at the end of the calendar month prior to the week used to compute $Return_{i,t}$. We compute order imbalance measures on a stock-day basis and, for the regressions presented in Panels A, B, and D, average those daily correlations on a weekly basis over the 104-week sample period of August 3, 2020 to July 26, 2022. For any given stock-week observation, we eliminate all that are missing any of the eight (four based on the wholesaler(s)'s retail orders and four based on BJZZ's methodology of identifying the same set of stocks in TAQ) or stock-weeks with any extreme order imbalance values (0, +1, or -1) as all of these indicate that there is a paucity of observations when computing the order imbalance measures. We separate securities into quintiles on a monthly based on retail trading intensity defined as the sum of a security's executed marketable order flow from the six largest wholesalers' SEC 605 Reports divided by that security's total share volume for that month from TAQ. Securities are placed into quintiles with a one-month lag. The regressions in Panel A (B) average 1,604 (634) securities. We run Fama-Macbeth regressions with Newey West standard errors. We report the results of the regression in Table 12 for Newey West lag = 0 as additional lags do not change our conclusions. We report only the coefficient estimates for the lagged order imbalance measures to conserve space. *, **, and *** indicates statistical significance at the 0.10, 0.05, and 0.01 level, respectively.

Panel A. All securities with at least 10,000 shares traded in the month. Quintiles are formed by ranking stocks by the percentage of average daily trading volume that is retail. Quintile 1 has stocks where retail trading comprises less than 4.22% of average daily trading volume and Quintile 5 contains stocks where retail trading makes up more than 37.64% of average daily trading volume.

	MROIBTRD		MROIBVOL		ODDMROIBTRD		ODDMROIBVOL		Retail Intensity Breakpoint
Retail Intensity	BJZZ	Actual	BJZZ	Actual	BJZZ	Actual	BJZZ	Actual	
Quintile 1 (Low)	-0.0008 (-0.67)	0.0002*** (3.12)	0.0005 (0.73)	0.0020*** (3.13)	-0.0009 (-0.85)	0.0011 (1.38)	-0.0001 (-0.09)	0.0011 (1.50)	4.22%
Quintile 2	0.0016 (1.27)	0.0025*** (3.32)	0.0013 (1.38)	0.0025*** (3.37)	0.0013 (1.18)	0.0007 (0.71)	0.0014 (1.42)	0.0006 (0.69)	10.78%
Quintile 3	-0.0013 (-0.74)	0.0043*** (3.75)	0.0016 (1.21)	0.0041*** (3.62)	-0.0025 (-1.45)	0.0020 (1.31)	-0.0009 (-0.65)	0.0019 (1.26)	23.69%
Quintile 4	0.0062** (2.53)	0.0020 (1.27)	0.0052** (2.28)	0.0017 (1.16)	0.0023 (0.96)	0.0016 (0.74)	0.0022 (1.16)	0.0017 (0.75)	37.64%
Quintile 5 (High)	0.0107*** (3.51)	0.0040* (1.82)	0.0068** (2.40)	0.0042* (1.88)	0.0091*** (2.74)	0.0062** (2.46)	0.0061** (2.51)	0.0061** (2.45)	

Panel B. Common stocks traded on multiple exchange (i.e., BJZZ sample restrictions). All securities with at least 10,000 shares traded in the month. Quintiles are formed by ranking stocks by the percentage of average daily trading volume that is retail. Quintile 1 has stocks where retail trading comprises less than 2.86% of average daily trading volume and Quintile 5 contains stocks where retail trading makes up more than 17.81% of average daily trading volume.

	MROIBTRD		MROIBVOL		ODDMROIBTRD		ODDMROIBVOL		Retail Intensity Breakpoint
Retail Intensity	BJZZ	Actual	BJZZ	Actual	BJZZ	Actual	BJZZ	Actual	
Quintile 1 (Low)	-0.0008	0.0009	-0.0004	0.0009	0.0001	0.0005	0.0000	0.0005	
	(-0.46)	(1.05)	(-0.33)	(1.10)	(0.04)	(0.53)	(0.03)	(0.56)	2.86%
Quintile 2	-0.0007	0.0029***	0.0015	0.0028***	-0.0010	0.0019	-0.0001	0.0019	
	(-0.41)	(3.06)	(1.18)	(3.07)	(-0.63)	(1.64)	(-0.03)	(1.63)	4.68%
Quintile 3	0.0022	0.0027***	0.0004	0.0027***	-0.0026**	0.0016	0.0026**	0.0016	
	(1.31)	(2.92)	(0.27)	(2.86)	(2.03)	(1.57)	(2.06)	(1.56)	8.32%
Quintile 4	0.0010	0.0052***	0.0018	0.0049***	-0.0007	0.0029**	0.0014	0.0028**	
	(2.53)	(1.27)	(1.28)	(4.67)	(-0.37)	(2.12)	(0.96)	(2.07)	17.81%
Quintile 5 (High)	0.0075***	0.0045*	0.0049*	0.0043***	0.0057**	0.0033*	0.0044**	0.0033*	
	(3.20)	(2.87)	(1.91)	(2.71)	(2.42)	(1.76)	(2.09)	(1.76)	